A Neural Network Model of Concept-influenced Segmentation

Robert L. Goldstone (**rgoldsto@indiana.edu**)
Department of Psychology, Indiana University
Bloomington, IN 47405 USA

Abstract

Several models of categorization assume that fixed perceptual representations are combined together to determine categorizations. This research explores the possibility that categorization experience alters, rather than simply uses, descriptions of objects. Based on results from human experiments, a model is presented in which a competitive learning network is first given categorization training, and then is given a subsequent segmentation task, using the same network weights. Category learning establishes detectors for stimulus parts that are diagnostic, and these detectors, once established, bias the interpretation of subsequent objects to be segmented.

Concept Learning and Perception

The current research explores the influence that learning a new concept has on the segmentation of objects into component parts. Recently a number of researchers have argued that in many situations, concept learning influences the featural descriptions used to describe a set of objects. Rather than viewing perceptual descriptions as fixed by low-level sensory processes, this view maintains that perceptual descriptions are dependent on the higher-level processes that use the descriptions (Goldstone, Steyvers Spencer-Smith, & Kersten, 2000; Schyns, Goldstone, & Thibaut, 1998). Evidence for this view comes from the study of expert/novice differences (Lesgold et al., 1988), influences of acquired concepts on the interpretation of stimuli (Wisniewski & Medin, 1994), and influences of category learning on psychophysical measurements of perceptual sensitivity (Goldstone, 1994).

Experiential Influences on Object Segmentation

One type of influence of concept learning on perceptual learning may be to alter how objects are segmented into parts. Objects often have more than one possible segmentation. The letter "X" can be viewed as comprised of two crossing diagonal lines, or as a "V" and an upside-down "V" that barely touch at their vertices. The segmentation of scenes into parts depends upon experience. Behrmann, Zemel, and Mozer (1998) found that judgments about whether two parts had the same number of humps were faster when the two parts belonged to the same object rather than different objects. Further work has found an influence of experience on subsequent part comparisons. Two stimulus components are interpreted as belonging to the same object if they have co-occurred many times (Zemel, Behrmann, Mozer, & Bavelier, 1999). Thus, experience with particular feature combinations determines whether or not features will be integrated into a single object.

Pevtzow and Goldstone (1994; reported in Goldstone et al., 2000) explored the influence of category learning on segmentation with the materials shown in Figure 1. We pursued the idea that how psychologically natural a part is might depend on whether it has been useful for previous categorizations. Naturalness was measured by how quickly subjects could confirm that the part was contained within a whole object (Palmer, 1978). To test this conjecture, we gave participants a categorization task, followed by part/whole judgments. During categorization, participants were shown distortions of the objects A, B, C, and D shown in Figure 1. The objects were distorted by adding a random line segment that connected to the five segments already present. Subjects were given extended training with either a vertical or horizontal categorization rule. For participants who learned that A and C were in one category, and B and D were in another (a vertical categorization rule) the two component parts at the bottom of Figure 1 were diagnostic. For participants who learned that A and B belonged in one category, and C and D belonged to the other category (a horizontal rule), the components on the right were diagnostic.

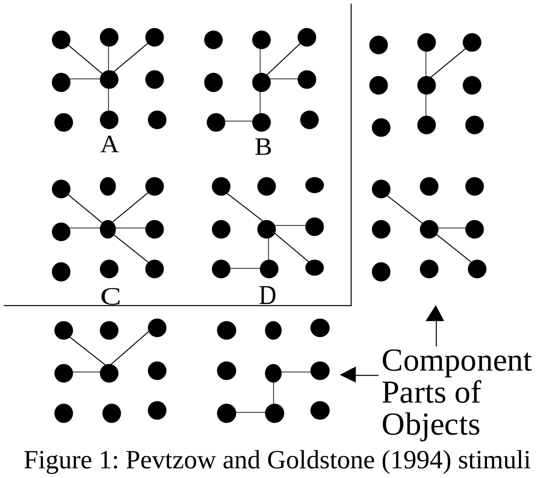

Figure 1: Pevtzow and Goldstone (1994) stimuli

During part/whole judgments, participants were shown a whole, and then a part, and were asked whether the part was contained in the whole. Participants were given both present and absent judgments, and examples of these judgments are shown in Figure 2. Note that the two parts shown in Figure 2 were both potentially diagnostic during the earlier categorization training. Whether or not a part was diagnostic was independent of the appearance of the part itself, depending only on how the four objects of Figure 1 were grouped into categories.

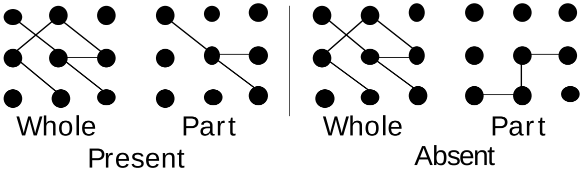

Figure 2: Part/whole present and absent judgments

The major result was that subjects were faster to correctly respond “present” when the part was diagnostic than when it was non-diagnostic. To the extent that one can find response time analogs of signal detection theory sensitivity and bias, this effect seems to be a sensitivity difference rather than a bias difference, because absent judgments also tended to be faster for diagnostic than nondiagnostic parts. Given that a category part that was diagnostic for the horizontal categorization group was nondiagnostic for the vertical group, these results indicate that it is not simply the physical stimulus properties that determine how readily a person can segment an object into a particular set of components; segmentation is also influenced by the learned categorical diagnosticity of the components

Modeling Interactions Between Concept Learning and Segmentation

We model the result from these experiment using a modified competitive learning network (Rumelhart & Zipser, 1985). As with the experiment, the network is first given categorization training, and then is given a subsequent segmentation task, using the same network weights. The goal of the modeling is to show how categorization training can prime the segmentation network such that objects will tend to be segmented into parts that were previously diagnostic for categorization.

The Categorization Network

The categorization network has three layers of units: one representing the input patterns, one representing a bank of learned detectors, and one reflecting the category assignments of the inputs. Both the weights from the input patterns to the detectors and the weights from the detectors to categories are learned. The categorization task uses a modified unsupervised competitive learning algorithm (Rumelhart & Zipser, 1985), but includes a top-down influence of category labels that incorporates supervised learning. The network begins with random weights from a two-dimensional input array to a set of detector units, and from the detectors to the category units. When an input pattern is presented, the unit with the weight vector that is closest to the input pattern is the "winner," and will selectively adjust its weights to become even more specialized toward the input. By this mechanism, the originally homogenous detectors will become differentiated over time, splitting the input patterns into categories represented by the detectors. The competitive learning algorithm automatically learns to group input patterns into the clusters that the patterns naturally form. However, given that we want the detectors to reflect the experiment-supplied categories, we need to modify the standard unsupervised algorithm. This is done by including a mechanism such that detectors that are useful for categorizing an input pattern become more likely to win the competition to learn the pattern. The usefulness of a detector is assumed to be directly proportional to the weight from the detector to the presented category which is provided as a label associated with an input pattern. The input-to-detector weights do not have to be set before the weights from detectors to categories are learned.

In addition to modifying the unsupervised development of hidden-layer detectors by considering their usefulness for categorization, a second modification of the standard competitive learning algorithm is required to fix one of its general problems in optimally making use of all detectors to cover a set of input patterns. This problem is that if multiple input patterns are presented that are fairly similar to each other, there will be a tendency for one detector to be the winner for all of the patterns. As a result, the winning detector’s weight vector will eventually become similar to the average of the input patterns’ activations, while the rest of the detectors do not learn at all. This situation is suboptimal because the input patterns are not covered as well as they would be if the unchanging detectors learned something. The standard solution to this problem is called "leaky learning" and involves adjusting both winning and losing detectors, but adjusting losing detectors at a slower rate (Rumelhart & Zipser, 1985). To understand the more subtle problem with this solution, imagine, for example, that four input patterns naturally fall into two groups based on their similarities, and the network is given four detectors. Ideally, each of the detectors would become specialized for one of the input patterns. However, under leaky learning, one detector will tend to become specialized for one cluster, a second will become specialized for the other cluster, and the remaining two detectors will be pulled equally by both clusters, becoming specialized for neither. Note that it does not help to supplement leaky learning by the rule that the closer a detector is to an input pattern, the higher its learning rate should be. There is no guarantee that the two "losing" units will evenly split such that each is closer to a different cluster.

Other researchers have noted related but not identical problems with competitive learning and have suggested solutions (Grossberg, 1987). Our current solution is to conceptualize competitive learning as not simply a competition among detectors to accommodate a presented input pattern, but also as a competition among input patterns to be accommodated by a given detector. Input patterns are presented sequentially to the network, and as they are presented, the closest input pattern to each detector is determined. The learning rate for a detector is set to a higher value for its closest input pattern than for other inputs. In this manner, detectors that are not the winning detector for a pattern can still become specialized by becoming unequally influenced by different patterns. In addition, the learning rate for a detector when presented with an input pattern will depend upon how well the input is currently covered by existing detectors. This dependency is required to allocate detectors to input regions where they are required. Putting these considerations together, the activation of detector i when presented with pattern p, is

$$A_{i,p} = \sum_{h=1}^{n} I_{h,p} W_{i,h} + \sum_{j=1}^{c} STW_{j,i}$$

where $I_{h,p}$ is the activation of input unit h for pattern p, $W_{i,h}$ is the weight from input h to detector i, S is the strength of the top-down pressure on detector development, T is the teacher signal (if Pattern p belongs to Category j then T=1, otherwise T=-1), and $W_{j,i}$ is the weight from Detector i to Category Unit j. The second term increases the activation of a detector to the extent that it is useful for predicting the input pattern's categorization. The detector activation will determine which detector is the "winner" for an input pattern. As such, detectors that are useful for categorization will tend to become winners, thus increasing their learning rate.

Input-to-detector weights are learned via top-down biased competitive learning using the following equation for changing weights from input pattern h to Detector i:

$$\Delta W_{i,h} = \begin{cases} M(I_{h,p} - W_{i,h}) \text{ if } \forall x(A_{i,p} \geq A_{x,p}) \\ \begin{Bmatrix} N(I_{h,p} - W_{i,h})K_p \text{ if } \forall y(A_{i,p} \geq A_{i,y}) \\ O(I_{h,p} - W_{i,h})K_p \text{ otherwise} \end{Bmatrix} \text{otherwise} \end{cases}$$

where M, N, and O are learning rates (M>N>O), and K_p is the distance between pattern p and its closest detector. This distance is inversely related to the cosine of the angle between the vector associated with the closest detector and p. This set of learning rules may appear non-local in that all detectors are influenced by the closest detector to a pattern, and depend on previous presented inputs. However, the rules can be interpreted as local if the pattern itself transmits a signal to detectors revealing how well covered it is, and if detectors have memories for previously attained matches to patterns. When an input pattern is presented, it will first activate the hidden layer of detectors, and then these detectors will cause the category units to become activated. The activation of the category unit A_j will be

$$A_j = \sum_{i=1}^{d} A_i W_{j,i}$$

where d is the number of detectors. Detector-to-category weights are learned via the delta rule $\Delta W_{j,i} = L(T - A_j)A_i$ where L is a learning rate and T is the teacher signal described above.

We formed a network with 2 detectors units and 2 category units, and presented it with four input patterns. We gave the network four patterns that were used in experiments with human subjects. These patterns are not identical to the patterns shown in Figure 1, but are of the same abstract construction. When the patterns were categorized as shown in Figure 3A, such that the first two patterns belonged to Category 1, and the second two patterns belonged to Category 2, then on virtually every run, the detectors that emerged were those reflecting the diagnostic segments -- those segments that were reliably present on Category 1 or Category 2 trials. The picture within a detector unit in Figure 3 reflects the entire weight vector from the 15 X 15 input array to the detector. When the same patterns are presented, but are categorized in the orthogonal manner shown in Figure 3B, then different detectors emerge that again reflect the category-diagnostic segments. In both cases, each detector will have a strong association to one and only one of the category units. This is expected given that one of the factors influencing the development of detectors was their categorical diagnosticity. For the results shown here, and the later simulations to be reported, the following parameter values were chosen: M=0.1, N=0.05, O=0.02, and S=0.1. Activation values were between –1 and +1. One hundred passes through the input materials were presented to the network. In the example shown in Figure 3, only 30 passes with each of the 4 patterns was required for the complete specialization of detectors to input patterns.

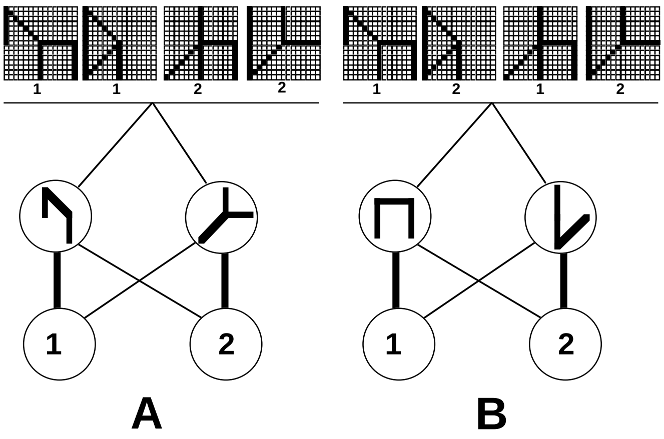

Figure 3: Categorization-dependent detectors are acquired

The Segmentation Network

The basic insight connecting categorization and segmentation tasks is that segmentation can also be modeled using competitive learning, and thus the two tasks can share the same network weights and consequently influence on each other. Competitive learning for categorization sorts complete, whole input patterns into separate groups. Competitive learning for segmentation takes a single input pattern, and sorts the pieces of the pattern into separate groups. For segmentation, instead of providing a whole pattern at once, we feed in the pattern one pixel at a time, so instead of grouping patterns, the network groups pixels together. Thus, each detector will compete to cover individual pixels of an input pattern such that the detector with the pixel-to-detector weight that is closest to the pixel's actual value will adapt its weight toward the pixel's value, and inhibit other detectors from so adapting. With this technique, if the pattern in Figure 4 is presented to the network, the network might segment it in the fashion shown in Panel A. Panels A-D show the weights from the 15 X 15 input array to each of two detectors, and reflect the specializations of the detectors. The two segments are complements of each other — if one detector becomes specialized for a pixel, the other detector does not.

Unfortunately, this segmentation is psychologically implausible. No person would decompose the original figure into these parts. To create psychologically plausible segmentations, we modify the determination of winners. Topological constraints on detector creation are incorporated by two mechanisms: A) input-to-detector weights "leak" to their neighbors in an amount proportional to their proximity in the 15 X 15 array, and B) input-to-detector weights also spread to each other as a function of their orientation similarity, defined by the inner-product of four orientation filters The first mechanism produces detectors that tend to respond to cohesive, contiguous regions of an input. The second mechanism produces detectors that follow the principle of good continuation, dividing the figure "X" into two crossing lines rather than two kissing sideways "V"s, because the two halves of a diagonal line will be linked by their common orientation. Thus, if a detector wins for pixel X (meaning that the detector receives the more activation when Pixel X is on than any other detector), then the detector will also tend to handle pixels that are close to, and have similar orientations to, Pixel X. The segmentation network, augmented by spreading weights according to spatial and orientational similarity, produces segmentations such as the one shown in Panel B of Figure 4.

Although the segmentation in Panel B is clearly superior to Panel A's segmentation, it is still problematic. The pixels are now coherently organized in line segments, but the line segments are not coherently organized into connected parts. Spreading weights according to spatial similarity should ideally create segmentations with connected lines, but such segmentations are often not found because of local minima in the harmony function (the value N defined on the next page). Local minima occur when a detector develops specializations for distantly related pixels, and these specializations develop into local regions of mutually supporting pixels. Adjacent regions will frequently be controlled by different detectors. Each of the detectors may have sufficiently strong specializations for local regions that they will not be likely to lose their specialization, due to the local relations of mutual support.

Our solution to local minima is to incorporate simulated annealing, by which randomness is injected into the system, and the amount of randomness decreases as a function of time. Unlike standard annealing techniques, we reduce the amount of randomness in the system over time, but do so by basing the amount of randomness on the current structural goodness of a solution (Hofstadter & Mitchell, 1994).

The segmentation network works by fully connecting a 15 X 15 input array of pixel values to a set of N detectors. Although ideally the value of N would be dynamically determined by the input pattern itself, in the current modeling, we assume that each object is to be segmented into two parts (as did Palmer, 1978). When an input pattern is presented, the pixels within it are presented in a random sequence to the detectors, and the activation of Detector i which results from presenting Pixel p is

$$A_{i,p} = \sum_{h=1}^{n} I_h W_{i,h} S_{h,p}$$

where I_h is the activation of Pixel h, $W_{i,h}$ is the weight from Pixel h to Detector i, and S is the similarity between pixels h and p. As such, detectors are not only activated directly by presented pixels, but are also activated indirectly by pixels that are similar to the presented pixels. Thus, a detector will be likely to be strongly activated by a certain pixel if it is already activated by other pixels similar to this pixel.

The similarity between two pixels h and p is determined by

$$S_{h,p} = T \frac{\sum_{i=1}^{n} G_{ih} G_{ip} L_{i,h,p}}{n} + Ue^{-D_{h,p} C}$$

Where T and U are weighting factors, G_{ih} is the response of orientation filter i to Pixel h, $L_{i,h,p}$ is the degree to which Pixels h and p fall on a single line with an orientation specified by filter i, $D_{h,p}$ is the Euclidean distance between Pixels h and p, and C is a constant that determines the steepness of the distance function. Four orientation filters were applied, at 0, 45, 90, and 135 degrees. The response of each filter was

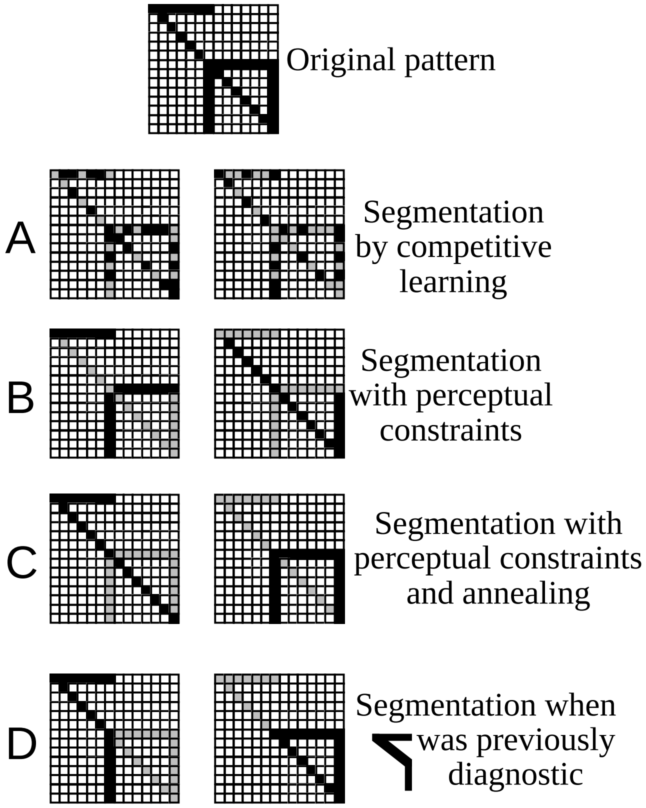
Original pattern
A
Segmentation by competitive learning
B
Segmentation with perceptual constraints
C
Segmentation with perceptual constraints and annealing
D
Segmentation when
was previously diagnostic

Figure 4. Segmentations of the original figure with incremental improvements from A-D.

found by finding the inner product of the image centered around a pixel and a 5 X 5 window with the image of one of the four lines. Thus, the greater the overlap between the line and the image, the greater will be the output of the filter for the line. The alignment of two pixels along a certain direction was found by measuring the displacement, in pixels, between the infinite length lines established by the two pixel/orientation pairs.

Pixel-to-Detector weights are learned via competitive learning:

$$\Delta W_{i,p} = \begin{cases} M(I_p - W_{i,p}) + Random(-N,+N) \text{ if } \forall x(A_{i,p} \geq A_{x,p}) \\ Random(-N,+N) \text{ otherwise} \end{cases}$$

Where M is a learning rate, and Random(-N,+N) generates Gaussian random noise between + and – N. The amount of noise, N, in adjusting weights is a function of the harmony across all detectors relative to R, the maximal harmony in the system:

$$N = R - \sum_{i=1}^{n}\sum_{p=1}^{m}\sum_{h=1}^{m} I_h I_p W_{i,h} W_{i,p} S_{h,p}$$

As such, if similar pixels in similar states have similar weights to detectors, then the harmony in the system will be high, and the amount of noise will be low. Thus, the amount of randomness in the weight learning process will be inversely proportional to the coherency of the current segmentation. These learning equations allow the network to regularly create the segmentation shown in Panel C of Figure 4.

In the simulations of the segmentation network to be reported, no attempt was made to find optimally fitting values of the constants. T and U were set at 0.5, M was set at 0.1, and C was set to 1.

Combining the Networks

Considered separately, the categorization and segmentation networks each can be considered to be models of their respective tasks. However, they were also designed to interact, with the aim of accounting for the results from Pevtzow and Goldstone's (1994) experiments with human subjects. The segmentation network, because it shares the same input-to-detector weights that were used for the categorization network, can be influenced by previous category learning. Detectors that were diagnostic for categorization will be more likely used to segment a pattern because they have already been primed. Thus, if a particular shape is diagnostic and reasonably natural, the network will segment the whole into this shape most of the time, as shown in Panel D Figure 4. In short, category learning can alter the perceived organization of an object. By establishing multi-segment features along a bank of detectors, the segmentation network is biased to parse objects in terms of these features. Thus, two separate cognitive tasks can be viewed as mutually constraining self-organization processes. Categorization can be understood in terms of the specialization of perceptual detectors for particular input patterns, where the specialization is influenced by the diagnosticity of a segment for categorization. Object segmentation can be viewed as the specialization of detectors for particular parts within a single input pattern. Object segmentation can isolate single parts of an input pattern that are potentially useful for categorization, and categorization can suggest possible ways of parsing an object that would not otherwise have been considered.

In order to model the results from the earlier human experiments, the network was first trained on distortions of the patterns A, B, C, and D shown in Figure 1, with either a horizontal or vertical categorization rule. As with the human experiment, the distortions were obtained by adding one random line segment to each pattern in a manner that resulted in a fully contiguous form. Following 30 randomly ordered presentations of distortions of the four patterns, the segmentation network was then presented with the original object shown in Figure 5. Segmentations were determined by examining the stable input-to-detector weight matrix for each of the two detector units.

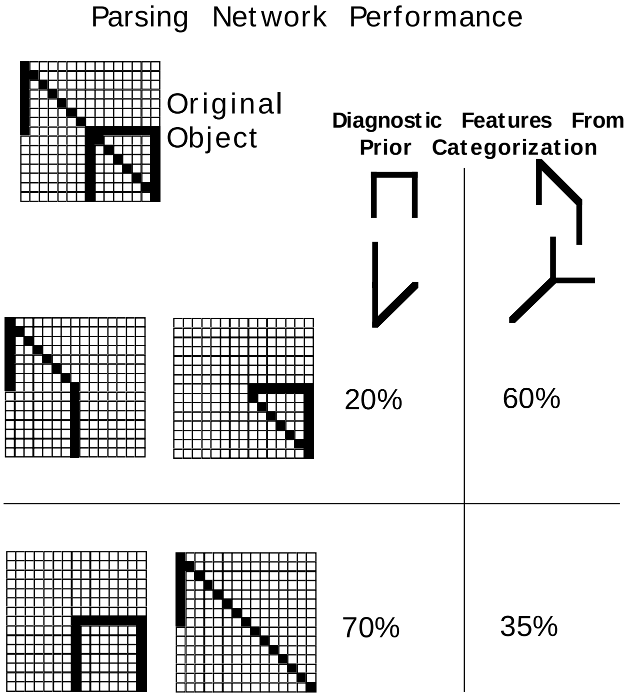
Parsing Network Performance
Original Object
Diagnostic Features From Prior Categorization
20%
60%
70%
35%

Figure 5. The segmentation of an ambiguous object is influenced by prior category learning.

One hundred subjects were simulated in each of the two pre-segmentation categorization conditions. As the results from Figure 5 indicate, the segmentation of the ambiguous original object is influenced by category learning. In particular, the original object tends to be segmented into parts that were previously relevant during category learning (column percentages do not add up to 100% because of rarely occurring alternative segmentations). As such, the results from Pevtzow and Goldstone (1994) are predicted under the additional assumption that response times in a part/whole task are related to the likelihood of generating a segmentation that includes the probed part.

In a subsequent test of the networks, the actual wholes used by Pevtzow and Goldstone (1994) in their part/whole task were presented to the segmentation network. Each whole was presented 200 times, 100 times preceded by each of the two possible categorization rules. Out of the 24 whole objects tested, segmentations involving categorization-relevant parts were produced more often than segmentations involving irrelevant parts for 19 of the objects. This comparison controls for any intrinsic differences in naturalness between segmentations of a whole object because the parts that are categorization-relevant for half of the simulated subjects are irrelevant for the other half. As such, the results from Figure 5 generalize to the actual materials used in the experiment. Human subjects and the simulation were exposed to same image-based materials, rather than presenting a digested and abstracted stimulus representation to the simulation.

Conclusions

A pair of neural networks were presented that learned to group multiple objects into categories, and learned to group parts from a single object into segments. More importantly, the computational modeling provides a mechanism by which one type of grouping influences the other. Category learning causes detectors to develop, and once these detectors have developed, there is a tendency to use the detectors when segmenting an object into parts.

Future work will be necessary compare the model to other existing models that allow for experience-dependent visual object segmentation (e.g. Behrmann et al., 1998; Mozer, Zemel, Behrmann, & Williams, 1992). Two extensions of the model would clearly be desirable: 1) allowing the model to determine for itself how many segments a pattern should be decomposed into, and 2) allowing the computed segmentation of a single pattern to influence its categorization. The latter extension is required to fit human experimental evidence suggesting that not only does category learning influence segmentation, but the perceived segmentation of an object influences its categorization (Schyns et al, 1998; Wisniewski & Medin, 1994).

The computational model, and associated experimental results, support theories that propose that categorization does not simply employ fixed descriptions such as geons, textons, holons, oriented lines segments, or spatial filters, but also creates new object descriptions. The primary advantage of such a state of affairs is that the perceptual system can become tuned and specialized to environmental demands. Cognitive science researchers who have proposed particular fixed sets of primitives have been clever, and have designed primitives that are useful for representing words, objects, and events. However, everyday people may be almost as clever as these researchers have been, and may be able to come up with their own sets of elements tailored to important categorizations (Schyns et al, 1998). Once created, these elements are then used for interpreting subsequently encountered objects. To the person who has a hammer, the world looks like a nail, and to the person who has learned that a particular configuration is relevant for categorization, the world looks like it is composed out of that configuration.

Acknowledgments

This research was funded by National Institute of Health Grant R01MH56871-01A2, and a Gill fellowship.

References

Behrmann, M., Zemel, R. S., & Mozer, M. C. (1998). Object-based attention and occlusion: Evidence from normal participants and a computational model. Journal of Experimental Psychology: Human Perception and Performance, 24, 1011-1036.

Goldstone, R. L. (1994). influences of categorization on perceptual discrimination. Journal of Experimental Psychology: General, 123, 178-200.

Goldstone, R. L., Steyvers, M., Spencer-Smith, J., & Kersten, A. (2000). Interactions between perceptual and conceptual learning. in E. Diettrich & A. B. Markman (Eds.) Cognitive Dynamics: Conceptual Change in Humans and Machines. (pp. 191-228). Lawrence Erlbaum and Associates.

Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. Cognitive Science, 11, 23-63.

Hofstadter, D. R., & Mitchell, M. (1994). The Copycat project: A model of mental fluidity and analogy-making. In K. J. Holyoak and J. A. Barnden (Eds.) Advances in Connectionist and Neural Computation Theory, Volume 2. (pp. 31-112). Norwood, NJ: Ablex.

Lesgold, A., Glaser, R., Rubinson, H., Klopfer, D., Feltovich, P., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), The nature of expertise. (pp. 315-335). Hillsdale, NJ: Erlbaum.

Mozer, M. C., Zemel, R. S., Behrmann, M., & Williams, C. K. I. (1992). Learning to segment images using dynamic feature binding. Neural Computation, 4, 650-665.

Palmer, S. E. (1978). Structural aspects of visual similarity. Memory & Cognition, 6, 91-97.

Pevtzow, R., & Goldstone, R. L. (1994). Categorization and the parsing of objects. Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society. (pp. 717-722). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. Cognitive Science, 9, 75-112.

Schyns, P. G., Goldstone, R. L, & Thibaut, J. (1998). Development of features in object concepts. Behavioral and Brain Sciences, 21, 1-54.

Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. Cognitive Science, 18, 221-281.

Zemel, R. S., Behrmann, M., Mozer, M. C., & Bavelier, D. (1999). Experience-dependent perceptual grouping and object-based attention. Unpublished manuscript.